CAP Protocol Common Data Elements
Rebecca Crowley
July 16, 2004

Use Case	**Developing standards for caTISSUE manual and automated annotation**
Case:	Although it's still not yet entirely decided as to the specific projects that will be undertaken by TBPT W/S – two ideas that are being seriously considered are (1) an inventory system, and (2) a system for annotation. Annotation can be considered to represent the addition of information to tissue resources that enables querying for subsets of resources, AND interpretation of any test performed on that resource. Annotation includes lots of things: information about disease process, organ, tissue, surgical procedure, amount of tissue in a block, procedure used to obtain the tissue, how the tissue was processed, what are the results of various ancillary tests (IPOX, Flow Cytometry), what happened to the patient over time, etc. etc.
Actors:	Actually there are several – pathologists who are manually annotating, tissue banking application developers, researchers who are creating protocols, research consortiums, vendors.
Background:	Any tissue bank that maintains these annotations of tissue generally does so manually, creating their own data elements, and maintaining them in their own data structures. Tissue Consortiums (CPCTR, etc) generally agree on a set of minimum DE and a common schema. We are interested in both supporting the manual annotation process, and also moving towards more automated methods of extracting information directly from text reports, and eventually from LIS when structured data is available via synoptic reporting. Also keep in mind that some projects (SPIN for example) are trying to alter the conventional idea of what a tissue bank is – SPIN suggests that the entire tissue archive is actually a Tissue Bank! Similar to ideas of "lightly" banked tissue. One element of annotating tissue is annotation of organ, tissue, disease, findings - at least some of which are usually represented in the Pathology Report. Pathology reports at most Cancer Centers share a fairly reproducible set of data elements (all of which are in free text ;-) that are based on standards specific to a given organ and procedure and

	diagnosis. For example, reports from a needle biopsy of prostate and a parathyroidectomy will have very different information – and this information is quite predictable at most university medical centers. The kinds of annotations we expect TB to want to apply to tissue in these two cases will overlap significantly (but not entirely) with the information in the Pathology Report.
Problem:	We want to develop a set of Common Data Elements that describes the usual information found in pathology reports. We anticipate that this could be used in several ways – (1) to act as a bank of standardized CDE's that will address the majority (but not the entirety) of disease/finding/organ annotations. It will not address many other types of annotation (tissue provenance for example). (2) To assist us in Information Extraction from free text reports, and (3) to act as an impetus (and resource) for commercial LIS vendors to incorporate caBIG standards so that we can actually capture this data in coded form from the start!
Solution:	We would like to use the CAP protocols to act as the basis for these CDE's. There are many issues with this, including: (1) how well do the disease categories, findings, units of measure, etc used in the CAP protocols map to existing terminologies: (Thesaurus, SNOMED) – where are additions to vocabularies needed? How much disambiguation is needed? (2) How well does ISO/IEC 11179 apply ? (3) What should the relationship with the original editors of the Protocols be? What about CAP? (4) How should we handle versioning of a set of CDE's? (4) How to tie these annotation CDEs into the information model?

Use Case 2	**Annotating to Adult Mouse Anatomy Vocabulary – more detail**
Case:	Annotate the location of an expression result in an adult mouse
Primary Actor:	Expression curator
Background:	The curator has a result in hand of the form: Assay type A using probe P detects expression of gene G in anatomical structure S in genotype (strain) T and needs to record this result.
Problem:	The curator 1) enters assay type A (ignore details) assay types are: Immunohistochemistry, RNA in situ, RNA in situ reporter (knock-in), Northern blot, Western blot, RNase protection, Nuclease S1, RT-PCR 2) probe P (ignore details) 3) enters gene G (ignore details) 4) enters genotype (strain) T (ignore details) 5) selects anatomical structure S from a browser (or types part S to get a list of possible structures and selects one) The system records the result.
Solution:	1) the anatomy must support different levels of spatial resolution to support the different assay types. Immunohistochemistry and RNA in situ can be very high resolution, Northern and Western blot are low resolution. (the point is to describe expression data at different levels of resolution and integrate them all) **2)** still doesn't tell us much about the vocabulary

Use Case 3	**Query using Adult Mouse Anatomy Vocabulary – 30K feet**
Case:	Query for expression results in an adult mouse
Primary Actor:	Researcher
Background:	Researcher wants to know genes are expressed in a specific anatomical structure (or substructure).
Problem:	Researcher selects a structure from an anatomy browser OR types some part of an anatomy term to get a list of possible structures and selects one. System returns all expression results annotated to that structure or any substructures. *Question: talk about other possible query parameters? Again, how much do we focus on the system vs. the vocabulary?*
Solution:	1) the anatomy must be hierarchical so that queries for structures can return results for substructures **2)** synonyms must be supported since the user can type anatomy terms

Use Case 4	**Annotating to Adult Mouse Anatomy Vocabulary – Sampling issue**
Case:	Annotate the location of an expression result in an adult mouse – where the structure is not very specific
Primary Actor:	Expression curator
Background:	The curator has a result in hand of the form: Assay type A using probe P detects expression of gene G in anatomical structure S in genotype (strain) T But structure S is not very specific. Examples: • just "lung" instead of "left lung" or "right lung" • just "alveolus" instead of "left lung alveolus" or "right lung alveolus" and needs to record this result.
Problem:	The curator (same above?) 6) enters assay type A (ignore details) assay types are: Immunohistochemistry, RNA in situ, RNA in situ reporter (knock-in), Northern blot, Western blot, RNase protection, Nuclease S1, RT-PCR 7) probe P (ignore details) 8) enters gene G (ignore details) 9) enters genotype (strain) T (ignore details) 10) selects anatomical structure S from a browser (or types part S to get a list of possible structures and selects one) The system records the result.
Solution:	1) the anatomy must support different levels of genericity. Curators need to be able to add more generic structures as needed.

Use Case 5	**Querying using Adult Mouse Anatomy Vocabulary – Multiple relationships**
Case:	Query for expression results in an adult mouse – but the same structure can be viewed as a substructure of different parents
Primary Actor:	Researcher
Background:	Researcher wants to know genes are expressed in a specific anatomical structure (or substructure).
Problem:	Researcher selects a structure from an anatomy browser, say "sensory organ system" Results annotated to "eye" should be returned. OR Researcher selects a structure from an anatomy browser, say "head" Again, results annotated to "eye" should be returned.
Solution:	**1)** the vocabulary must support a DAG structure.

Other Issues not addressed by the above Use Cases:

1) entering negative expression results
2) querying for negative expression results
3) updating the anatomy, say to add more detailed substructures or add more generic structures, adding synonyms, etc.
4) browsing the anatomy
5) where to define vocabulary boundaries, e.g., does the anatomy include cell types?
6) what to do about describing abnormal anatomy, e.g., extra digits
7) the anatomy should be used in phenotype descriptions (e.g., abnormal development of X), including cancer phenotypes

Our one concrete thought about all this:

1) Use cases that are intended to guide vocabulary development (or selection of existing vocabularies) should include as many actual examples as is feasible, particularly around obviously sticky areas.